融合背景估计与 U-Net 的文档图像二值化算法 *

熊 炜[a, b]，王鑫睿[a]，王 娟[a,b]，刘 敏[a, b]，曾春艳[a, b]
(湖北工业大学 a. 电气与电子工程学院; b. 太阳能高效利用湖北省协同创新中心，武汉 430068)

摘　要：针对低质量文档图像存在页面污渍、墨迹浸润、背景纹理等多种退化因素，提出一种融合背景估计与 U 型卷积神经网络(U-Net)的文档图像二值化算法。该算法首先进行图像对比度增强，然后通过形态学闭操作来估计文档图像背景，并利用全卷积网络，即 U-Net 对背景减除图像进行前景背景分割，最后采用全局最优阈值处理方法获得最终二值图像。实验结果表明，在 2016 和 2017 年国际文档图像二值化竞赛中本文算法的 *F* 值(*F*-measure，FM)、伪 *F* 值(pseudo *F*-measure，p-FM)、峰值信噪比(peak signal to noise ratio，PSNR)、距离倒数失真度量(distance reciprocal distortion，DRD)比性能次优的经典算法最高有 5.58%、2.47%、0.86dB、1.19%的性能提升。

关键词：文档图像二值化；对比度增强；形态学闭操作；U 型卷积神经网络；全局最优阈值处理
中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2018.07.0666

Document image binarization algorithm based on background estimation and u-net

Xiong Wei[a, b], Wang Xinrui[a], Wang Juan[a, b], Liu Min[a,b], Zeng Chunyan[a,b],
(*a. School of Electrical & Electronic Engineering, b. Hubei Collaborative Innovation Center for High-efficiency Utilization of Solar Energy, Hubei University of Technology, Wuhan* 430068, *China*)

Abstract: Degraded document images have various degradation factors, such as page stains, ink bleed-through, and background texture. We propose a novel document image binarization algorithm based on background estimation and U-Net. The algorithm first performs image contrast enhancement, and estimates the document background via morphological closing operations. We then adopt a fully convolutional network, namely the U-Net, to extract the foreground text from the document background. Finally, the global optimal thresholding method is used to obtain the resulting binary image. The proposed technique has been extensively evaluated over the recent DIBCO benchmark datasets. Experimental results show that our proposed method outperforms other state-of-the-art document image binarization algorithms in terms of F-measure, pseudo F-measure, PSNR and DRD, with 5.58%、2.47%、0.86dB and 1.19%.

Key words: document image binarization; contrast enhancement; morphological closing operation; u-net; global optimal thresholding

0　引言

图像二值化作为识别与分析文档图像的预处理步骤，被广泛应用于古籍文档修复，签名验证等领域[1]。受物理条件和人为因素的影响，低质量文档图像具有页面污渍等复杂背景特性[2]，导致图像的目标信息与背景差异较小，因此低质量文档图像二值化极具挑战性[3]。

文档图像二值化算法划分为全局阈值法，局部阈值法和混合阈值法[4]。全局阈值法通过选取固定阈值 G_T 作用于所有像素点，将图像分为前景与背景两大类，算法优点在于复杂度较低，但处理具有复杂背景特性的图像会造成文本丢失，经典的全局阈值法有 Otsu 算法[5]等。局部阈值法基于滑动窗与图像的卷积操作来设定每个像素点的阈值。Wolf[6]利用图像的归一化对比度和邻域内标准差与灰度均值，从而进行局部二值化。局部阈值法还包括 Sauvola 算法[7]、Niblack 算法[8]等。此类算法适用性强，但性能依赖邻域的窗口尺寸。

Su 等人[9]将局部灰度值作归一化处理，以抑制图像背景亮度不均匀产生的影响(LMM 算法)，但处理边缘区域时存在字符中空现象。Lu 等人[10]基于邻域内像素点的差异，检测出字符边缘并估计字符宽度进而完成二值化(BESE 算法)，但算法抑制背景污渍的能力较差。Howe 通过图割方法实现能量函数最小化完成图像分割[11]，并对算法的结构参数进行调优[12]，但是该类算法处理低对比度图像会损失部分笔画细节。Mesquita[13]提出一种人眼视觉模型，基于能量函数最小化与 I/F-Race 方法，区分文本与背景像素点。Kligler[14]依据文档图像的亮度变化将估计背景移除，利用图割算法得到二值结果，但算法会将浸润的墨迹错判为字符。Tensmeyer[15]结合相对暗特征与能量函数，采用 5 层全卷积神经网络完成图像二值化，该算法虽然能够较好的抑制背景污渍，但是容易造成笔画断裂。此外还有参数调优法[16]、分类器法[17]、聚类法[18]等混合阈值算法。本文将背景估计与 U 型卷积神经网络相结合，从而实现文档图像二值化，并通过实验对比分析验证本文方法的性能。

收稿日期：2018-07-26；**修回日期：**2018-09-25　**基金项目：**国家自然科学基金资助项目（61501178，61571182，61601177）；湖北省教育厅科学技术研究项目（D20161404）；太阳能高效利用湖北省协同创新中心开放基金资助项目（HBSKFZD2014011）

作者简介：熊炜（1976-），男，湖北宜昌人，副教授，博士，主要研究方向为机器人视觉、增强现实（xw@mail.hbut.edu.cn）；王鑫睿（1994-），男，湖北广水人，硕士研究生，主要研究方向为数字图像处理、图形图像；王娟（1983-），女，河北邯郸人，讲师，博士，主要研究方向为人工智能、图像处理；刘敏（1979-），女，河南周口人，副教授，博士，主要研究方向为计算机视觉、模式识别与智能系统；曾春艳（1986-），女，湖北武人，讲师，博士，主要研究方向为压缩感知、机器学习．

1　融合背景估计与 U 型卷积神经网络的文档图像二值化算法

1.1　算法流程

本文提出方法的模型如图 1 所示。

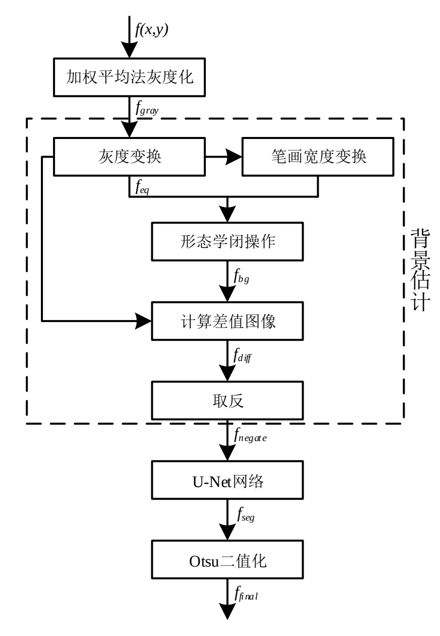


图 1　算法流程图

Fig. 1　Algorithm flowchart

a)对彩色文档图像 $f(x,y)$ 作灰度处理得到灰度图 f_{gray}，并对图像 f_{gray} 进行对比度增强；

b)利用形态学闭操作估计对比度增强图像 f_{eq} 的背景，其结构元的大小与文字笔画宽度有关；

c)结合对比度增强图 f_{eq} 与背景估计图 f_{bg}，得到去除背景图像 f_{negate}；

d)通过 U-Net 网络实现对背景减除图 f_{negate} 的分割得到图像 f_{seg}，并通过 Otsu 算法得到二值结果图像 f_{final}。

1.2　背景估计

对原始文档图像采用加权平均法，如式(1)所示，作灰度化处理得到图 f_{gray}，如图 2(b)所示。

$$f_{gray}(x,y)=0.299\,f_R(x,y)+0.587\,f_G(x,y)+0.114\,f_B(x,y) \tag{1}$$

其中：$f_R(x,y)$、$f_G(x,y)$、$f_B(x,y)$ 分别表示图像 R、G、B 三通道的分量。

对灰度图 f_{gray} 作线性灰度变换操作，如式(2)所示，得到图像 f_{eq} 如图 2(c)所示，此时图像 f_{eq} 中字符与背景之间的对比度得到提高。

$$f_{eq}(x,y)=\begin{cases} h_1 & f_{gray}(x,y)<l_1 \\ \dfrac{h_2-h_1}{l_2-l_1}\times(f_{gray}(x,y)-l_1)+h_1 & l_1<f_{gray}(x,y)<l_2 \\ h_2 & f_{gray}(x,y)>l_2 \end{cases} \tag{2}$$

其中：图像 f_{gray} 中灰度值小于 l_1 和大于 l_2 的像素点分别占整幅图像的 1%，$h_1=0$、$h_2=255$ 表示灰度变换图像 f_{eq} 的灰度值范围。

笔画宽度变换算法[19]利用 *Canny* 算子计算像素点 p 的梯度 d_p，沿射线 $r=p\pm n\cdot d_p(n>0)$ 寻找匹配点 q (梯度为 d_q)。若 d_p、d_q 的方向相反则计算欧式距离 $\|\overline{p-q}\|$，并将 p 与 q 之间的点赋值为 $\|\overline{p-q}\|$ (排除已被指定更小值的像素点)，若未能找到满足条件的像素点则将射线 $r=p\pm n\cdot d_p$ 舍弃，图像 f_{eq} 的笔画宽度估计 *SWE* 如式(3)所示。

$$SWE=\frac{1}{num}\sum_{s(x,y)\neq 0}s(x,y) \tag{3}$$

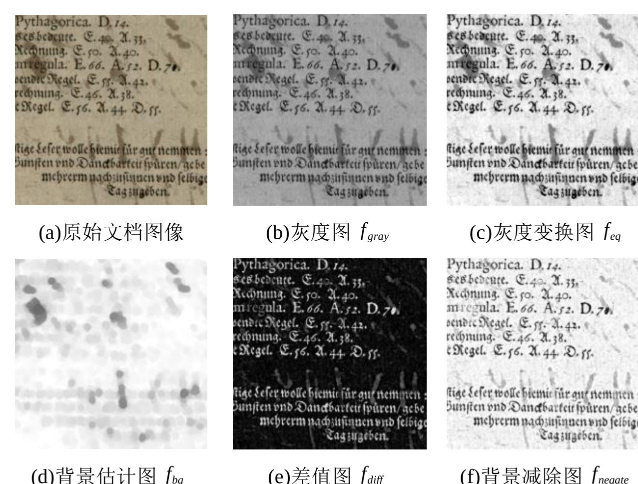
其中：*num* 表示笔画宽度变换算法的输出矩阵 $s(x,y)$ 中不为 0 的个数。

使用圆形结构元对灰度变换图 f_{eq} 进行形态学闭操作得到估计背景图 f_{bg}，如图 2(d)所示，结构元直径和笔画宽度有关，如式(4)所示。

$$f_{eq}\bullet b=(f_{eq}\oplus b)\ominus b \tag{4}$$

其中：$\oplus$ 表示膨胀，$\ominus$ 表示腐蚀，b 代表结构元，b 的直径为 $d=SWE+\Delta d$ (Δd 为增量)。

计算灰度变换图 f_{eq} 与背景估计图 f_{bg} 的绝对差值，得到图像 f_{diff}，如图 2(e)所示，并将差值图像 f_{diff} 取反获得背景减除图 f_{negate}，如图 2(f)所示。与灰度图像相比，抑制了图像背景污渍，有利于字符与背景的分割。



(a)原始文档图像　(b)灰度图 f_{gray}　(c)灰度变换图 f_{eq}

(d)背景估计图 f_{bg}　(e)差值图 f_{diff}　(f)背景减除图 f_{negate}

图 2　图像增强

Fig. 2　Image enhancement

为确定 Δd 的值，选取 2009 至 2014 年国际文档图像二值化竞赛(Document Image Binarization Contest, DIBCO)所提供的图像(共计 76 张)作为训练集，采用 Otsu 算法对背景减除图像进行二值化，并使用 F 值对输出结果进行评估，如式(5)所示，训练结果如表 1 所示。

$$\mathrm{FM}=\frac{2\times RC\times PR}{RC+PR} \tag{5}$$

其中：$RC=\dfrac{TP}{TP+FN}$ 代表查全率，$PR=\dfrac{TP}{TP+FP}$ 代表查准率；TP 表示为正确肯定像素数，FP 表示为错误肯定像素数，FN 表示为错误否定像素数。

表 1　训练结果

Table 1　Training results

Δd	FM	Δd	FM	Δd	FM
1	73.349	11	87.927	21	87.210
2	81.205	12	87.914	22	87.124
3	84.848	13	87.793	23	87.059
4	86.498	14	87.667	24	86.984
5	87.454	15	87.603	25	86.947
6	87.783	16	87.537	26	86.886
7	87.897	17	87.443	27	86.833
8	**88.070**	18	87.387	28	86.787
9	88.036	19	87.337	29	86.750
10	87.975	20	87.270	30	86.693

由表 1 可知 $\Delta d=8$ 时 F 值较大，说明此时背景抑制效果

较好，因此选取 Δd=8 进行对剩余数据集进行测试。

1.3　U 型卷积网络

1.3.1 U 型卷积神经网络模型

由于 U 型卷积神经网络(U-Net)[20]可以使用非常少的图像进行端到端训练，并且在 ISBI 神经元结构分割挑战赛中获得了优异的性能，因此本文采用 U-Net 对去除背景后的图像进行前景与背景分割，网络模型框如图 3 所示。

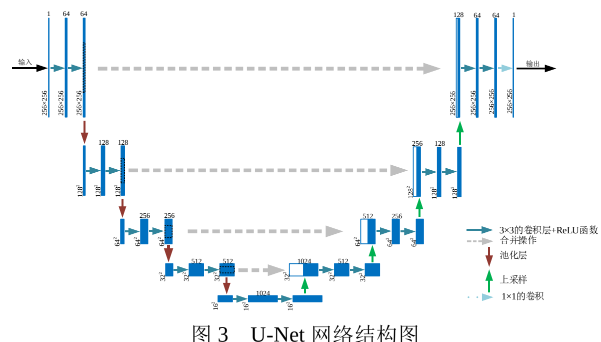


图 3　U-Net 网络结构图

Fig. 3　The architecture of U-Net

U-Net 架构主要由收缩路径和称扩张路径组成。在收缩路径中，基本单元由两个 3×3 卷积核及步长为 2 的 2×2 最大池化组成，使用修正线性单元作为激活函数，在下采样过程中得到低分辨率的高维特征图。在扩张路径中对高维特征图进行上采样，用 2×2 的上卷积核进行反卷积操作，特征通道数被减半，并与收缩路径中对应层级的特征图进行级联，采用与收缩路径中相同的两个 3×3 的卷积核进行卷积运算。最后一层中卷积核大小为 1×1，选择 Sigmoid 函数作为激活函数将输入特征向量映射为输出层，如式(6)所示。

$$S(\hat{y})=\frac{1}{1+e^{-\hat{y}}} \tag{6}$$

其中: $\hat{y}$ 为激活函数的输入特征，$S(\hat{y})\in(0,1)$ 为当前像素被识别为文本的概率值。

网络采用对数损失函数来反映预测值和真实值之间的差异，具体如式(7)所示，并通过反向传播算法对网络模型参数进行更新。

$$J_{CE}(y,\hat{y})=-\frac{1}{m}[\sum_{i=1}^{m}y_i\ln(S(\hat{y}_i))+(1-y_i)\ln(1-S(\hat{y}_i))] \tag{7}$$

其中: y_i 为真实值，$S(\hat{y}_i)$ 为预测值，m 为样本个数。

1.3.2 网络训练

选择经过背景估计操作后的文档图像及其标准图像作为网络的训练集，原始图像由 2009 至 2014 年 DIBCO 提供，考虑到图像大小不统一，其中最小图像高度为 263，因此实验选取步长为 214 的 256×256 窗口将图像裁切为相同大小的子图像，如图 4 所示。将 2027 张裁切图及其对应的标准图像作为输入对模型进行训练，并依据滑动窗位置完成对输出图像的拼接。

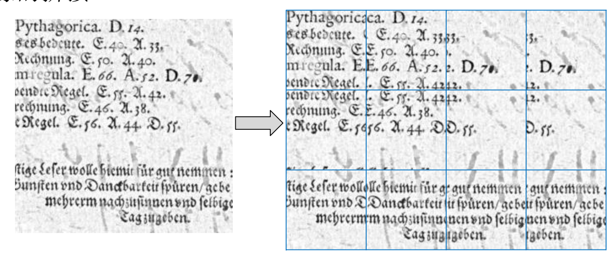


图 4　文档图像裁切

Fig. 4　Document image clipping

为最大限度利用显存，实验选取学习率为 10^{-4}，批次大小(Batch Size)为单张图像，网络训练的迭代次数为 10，网络训练的误差率变化如图 5 所示，误差率处于 0.013-0.014 说明模型达到稳定状态。

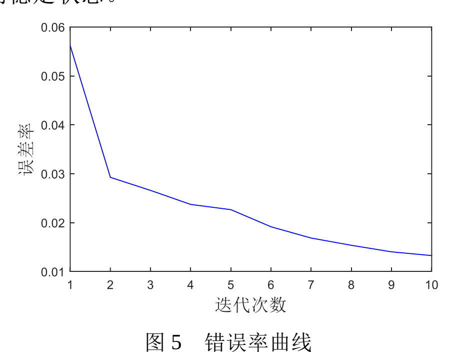
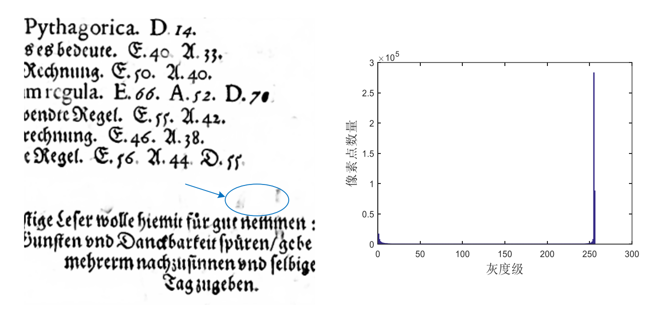


图 5　错误率曲线

Fig. 5　Error rate curve

1.4　全局最优阈值处理算法

U-Net 网络输出结果如图 6(a)所示，灰度直方图如图 6(b)所示。



(a)U-Net 网络输出图像　　(b)灰度直方图

图 6　U-Net 输出图像及其灰度直方图

Fig. 6　U-Net output and gray histogram

由图 6(b)可知 U-Net 的输出图像具有显著的双峰特性，因此本文选择全局最优阈值处理算法(Otsu)进行二值化，Otsu 算法根据图像的直方图分布特性得到 L 个灰度级分量 $p_i(i=0,1,2,\ldots,L-1)$，选择一个阈值 $k\in[0,L-1]$，分别计算前景像素所占比例 $P_1(k)=\sum_{i=0}^{k}p_i$ 与背景素所占比例 $P_2(k)=\sum_{i=k+1}^{L-1}p_i$。计算前景与背景像素点的灰度均值 $\mu_1(k)$、$\mu_2(k)$，如式(8)所示，并得到全局均值 $\mu=\sum_{i=0}^{L-1}ip_i$。

$$\begin{aligned}\mu_1(k)&=\frac{1}{P_1(k)}\sum_{i=0}^{k}ip_i\\ \mu_2(k)&=\frac{1}{P_2(k)}\sum_{i=k+1}^{L-1}ip_i\end{aligned} \tag{8}$$

类间方差作为可分性度量，如式(9)所示，根据类间方差最大原理，找到使得 $\sigma_{12}^2(k)$ 最大的 k，即最优阈值 k 对图像进行二值分类。

$$\begin{aligned}\sigma_{12}^2(k)&=P_1(k)[\mu_1(k)-\mu]^2+P_2(k)[\mu_2(k)-\mu]^2\\&=P_1(k)P_2(k)[\mu_1(k)-\mu_2(k)]^2\end{aligned} \tag{9}$$

最终二值图像如图 7(a)所示，图像的字符完整保留，图 6(a)中右下角污渍已被抑制，从视觉效果来看与标准图像较为接近。

2　实验与分析

实验测试集数据来源于 2016-2017 年国际文档图像二值化竞赛，共计 30 张图像。本文选取的评估指标为 F 值(*F*-measure, FM)、伪 F 值(pseudo F-measure, p-FM)、峰值信

噪比(PSNR)、距离倒数失真度量(DRD)，其中前三个指标值越大说明算法准确性越高，DRD 指标值越小说明像素分类错误率越低，具体定义请参考文献[21~23]。

(a)二值结果图像　　(b)标准图像

图 7　二值图像与对应标准图像

Fig. 7　Binary image and ground truth

表 2 是本文算法与 2016 和 2017 年国际文档二值化竞赛的前三名算法(分别表示为 TOP1、TOP2、TOP3)作对比，本文算法的四项评估指标值均优于前三名算法，说明本文提出的算法具有更高的准确性和鲁棒性。

表 2　2016—2017 年 DIBCO 前三名算法对比

Table 2　Comparison with the top three algorithms in the 2016-2017 DIBCO

竞赛	方法	FM	p-FM	PSNR	DRD
DIBCO2016	TOP1	87.61	91.28	18.11	5.21
	TOP2	88.72	91.84	18.45	3.86
	TOP3	88.47	91.71	18.29	3.93
	本文算法	**89.88**	**93.57**	**18.80**	**3.77**
DIBCO2017	TOP1	91.04	92.86	18.28	3.40
	TOP2	89.67	91.03	17.58	4.35
	TOP3	89.42	91.52	17.61	3.56
	本文算法	**91.64**	**93.33**	**18.40**	**3.33**

表 3 是本文算法与各类二值化算法在 2016 年 DIBCO 数据集上的输出结果，其中 – 代表没有该指标值，Time 代表时间复杂度，表示算法处理单张图像的平均速度(单位为 s)。除 Score 外表中参数取平均值，Score 值代表算法的综合性能，并根据 Score 值对算法排序，具体如式(10)所示。

$$\mathrm{Score} = \sum_{i=1}^{M}\sum_{j=1}^{N} R(i,j)$$
$$R = \begin{cases} \max(\psi_i(1:\beta))/\psi_i(j) & if \;\; 1 \le i \le 3 \\ \psi_i(j)/\min(\psi_i(1:\beta)) & if \;\; 4 \le i \le M \end{cases} \tag{10}$$

其中：$\psi = \{\mathrm{FM, p\text{-}FM, PSNR, DRD}\}$ 为所有评估指标的集合，N 为图像总数，M 为指标数，β 为单项指标总数。

为评估所有算法性能，本文在 NVIDIA GTX1080 8G GPU 下进行实验，由于未将时间复杂度 Time 纳入 Score 值的计算，算法执行效率 Time 不作为性能评估标准。

各类算法的评估结果中 Tensmeyer 算法有较大的 p-FM 值，说明 Tensmeyer 算法像素分类正确率较高。Otsu 算法选取固定阈值进行全局二值化，因此处理图像速度较快，本文算法有最高的 FM、PSNR 值，最小的 DRD 和 Score 值，由于复杂度较高因此处理图片速度较慢。

表 4 是 2017 年 DIBCO 数据集的测试结果，其中 Howe_base 算法有较大的 PSNR 值，说明 Howe_base 算法的输出结果与标准图像的相似度较高。本文方法有最高的 FM、p-FM 值，最小的 DRD 和 Score 值，从指标数值方面说明本文方法的综合性能比 Howe_alg3 等传统二值化算法更好，同时优于 Tensmeyer 所提出的神经网络模型。

表 3　DIBCO2016 各类算法评估指标值

Table 3　Evaluation results of various algorithms in DIBCO2016

算法	Rank	Score	FM	p-FM	PSNR	DRD	Time(s)
本文算法	1	**68.15**	**89.88**	93.57	**18.80**	**3.77**	6.763
Wolf[6]	2	72.45	87.75	91.20	18.49	4.16	0.076
Tensmeyer[15]	3	73.83	87.67	**94.73**	18.04	4.36	62.500
Howe_alg2[12]	4	74.30	88.12	92.76	18.30	4.38	45.455
Sauvola[8]	5	76.51	87.78	90.62	17.95	4.58	0.075
Howe_base[12]	6	76.58	87.81	92.59	18.03	4.62	1.812
Kligler[14]	7	82.16	87.61	92.40	18.11	5.21	43.478
Howe_alg3[12]	8	83.57	87.47	92.28	18.05	5.35	12.500
Howe_conf[11]	9	85.14	86.26	90.22	17.24	5.44	2.597
Otsu[5]	10	86.26	86.59	89.92	17.79	5.58	**0.015**
LMM[9]	11	86.95	84.75	88.94	17.64	5.64	0.086
Howe_alg1[12]	12	191.50	79.62	85.23	17.63	6.28	4.167
Niblack[7]	13	1035.85	42.31	42.56	6.94	106.08	0.074
BESE[10]*	14	-	87.53	91.33	18.62	4.10	16.949

表 4　DIBCO2017 各类算法评估指标值

Table 4　Evaluation results of various algorithms in DIBCO2017

算法	Rank	Score	FM	p-FM	PSNR	DRD	Time (s)
本文算法	1	**218.84**	**91.64**	**93.33**	18.40	**3.33**	7.811
Howe_alg3[12]	2	279.45	86.06	90.86	17.54	4.52	18.519
Kligler[14]	3	298.65	90.10	91.48	18.52	5.13	62.500
Howe_base[12]	4	308.61	89.87	91.19	**18.72**	5.36	2.033
Howe_alg1[12]	5	317.94	88.68	90.08	17.43	5.53	6.329
LMM[9]	6	319.42	85.86	87.24	18.39	5.60	0.093
Howe_conf[11]	7	334.42	82.25	87.79	15.84	5.78	2.833
Howe_alg2[12]	8	353.07	86.86	89.43	16.32	6.30	66.667
Otsu[5]	9	358.47	81.75	87.07	15.58	6.32	**0.019**
Tensmeyer[15]	10	366.38	87.29	89.40	16.89	6.62	71.429
Sauvola[8]	11	366.85	85.82	87.39	16.02	6.58	0.092
Wolf[6]	12	767.98	77.74	80.05	13.85	15.54	0.093
Niblack[7]	13	3054.91	47.76	48.35	7.16	66.86	0.092
BESE[10]*	14	-	89.57	91.53	17.03	4.55	19.608

注：*BESE 算法处理图片 DIBCO2016_8、DIBCO2017_1、DIBCO2017_2、DIBCO2017_3、DIBCO2017_4、DIBCO2017_5 出现报错，BESE 算法的指标值均为剔除报错图片后的结果，因此未计算 BESE 算法的 *Score* 值。

如图 8 所示，本文选取三张具有代表性的测试图像，分别为页面中脊图像、笔画纤细图像、墨迹浸润图像，并展示各类算法的二值化图像。

Otsu 算法将图像的页面中脊处与浸润的墨迹误判为前景；Wolf 算法处理笔画纤细图像和页面中脊图像时笔画断裂严重；Niblack 算法较好地保留了文字边界信息，但输出图像含有大量噪点；Sauvola 算法处理弱笔画时损失部分字符细节；LMM 算法在图像边缘处存在笔画中空现象；BESE 算法将浸润的笔墨错判为字符；文献[11，12]所提出的算法中，Howe_alg1 算法处理页面中脊和低对比度字符效果较好，Howe 的同类算法在墨迹浸润文档输出结果中仍有大量的背景被错判为前景；Kligler 算法二值化输出结果中也有笔画不连续现象，处理墨迹浸润图像能力较差；Tensmeyer 算法在抑制背景的同时，容易忽略细弱笔画。

本文提出的方法能较好地解决复杂背景的干扰问题，准确地分离出文本信息，因此从视觉效果方面体现本文算法的优势。



(a)原始文档图像

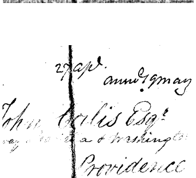
(b)Otsu 算法结果

(c)Wolf 算法结果





(d)Niblack 算法结果



(e)Sauvola 算法结果

(f)LMM 算法结果

(g)BESE 算法结果

(h)Howe_conf 算法结果

(i)Howe_base 算法结果

(j)Howe_alg1 算法结果

(k)Howe_alg2 算法结果

(l)Howe_alg3 算法结果

(m)Kligler 算法结果

(n)Tensmeyer 算法结果

(o)本文算法结果

图 8　各类算法二值化结果图

Fig. 8　Comparison of binarization results

3　结束语

本文提出了一种融合背景估计与U型卷积神经网络的文档图像二值化算法，通过背景估计操作抑制图像污渍背景，利用U-net网络将文档图像的像素进行分类，并结合Otsu算法完成精细二值分割。实验表明，在2016年和2017年国际文档图像二值化竞赛中本文算法的FM值、伪FM值、峰值信噪比(PSNR)、距离倒数失真度量(DRD)比性能次优的经典算法最高有5.58%、2.47%、0.86dB、1.19%的提升，充分证明本文算法的优越性能。

参考文献：

[1] Khitas M, Ziet L, Bouguezel S. Improved degraded document image binarization using median filter for background estimation [J]. Elektronika Ir Elektrotechnika, 2018, 24(3): 82-87.

[2] Sood R, Sharma M M. Binarization of ancient document images [J]. International Journal of Scientific Research and Engineering Studies, 2015, 2(6): 22-25.

[3] Milyaev S, Barinova O, Novikova T, *et al.* Fast and accurate scene text understanding with image binarization and off-the-shelf OCR [J]. International Journal on Document Analysis and Recognition, 2015, 18(2): 169-182.

[4] Eskenazi S, Gomez-Kreimer P, Ogier J M. A comprehensive survey of mostly textual document segmentation algorithms since 2008 [J]. Pattern Recognition, 2017, 64(1): 1-14.

[5] Otsu N. A threshold selection method from gray-level histograms [J]. IEEE Trans on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.

[6] Wolf C, Jolion J M, Chassaing F. Text localization, enhancement and binarization in multimedia documents [C]//Proc of the 16th International Conference on Pattern Recognition, Quebec: IEEE Press, 2002: 1037-1040.

[7] Niblack W. An introduction to digital image processing [M]. Englewood Cliffs: Prentice-Hall, 1986: 115-126.

[8] Sauvola J, Pietikeiinen M. Adaptive document image binarization [J]. Pattern Recognition, 2000, 33(2): 225-236.

[9] Su Bolan, Lu Shijian, Tan Chewlim. Binarization of historical document

images using the local maximum and minimum [C]//Proc of the 9th IAPR International Workshop on Document Analysis Systems.NewYork: ACM Press, 2010: 159-166.

[10] Lu Shijian, Su Bolan, Tan Chewlim. Document image binarization using background estimation and stroke edges [J]. International Journal on Document Analysis and Recognition, 2010, 13(4): 303-314.

[11] Howe N R. A laplacian energy for document binarization [C]//Proc of International Conference on Document Analysis and Recognition, Beijing: IEEE Press, 2011: 6-10.

[12] Howe N R. Document binarization with automatic parameter tuning [J]. International Journal on Document Analysis and Recognition, 2013, 16(3): 247-258.

[13] Mesquita R G, Silva R M, Mello C A, *et al*. Parameter tuning for document image binarization using a racing algorithm [J]. Expert Systems with Applications, 2015, 42(5): 2593-2603.

[14] Kligler N, Katz S, Tal A. Document enhancement using visibility detection [C]// Proc of the Conference on Computer Vision and Pattern Recognition, Salt: IEEE Press, 2018: 2374-2382.

[15] Tensmeyer C, Martinez T. Document image binarization with fully convolutional neural networks [J]. arXiv preprint arXiv: 170803276, 2017.

[16] Westphal F, Grahn H, Lavesson N. Efficient document image binarization using heterogeneous computing and parameter tuning [J]. International Journal on Document Analysis and Recognition, 2018, 21 (1): 41-58.

[17] Ahmadi E, Azimifar Z, Shams M, *et al.* Document image binarization using a discriminative structural classifier [J]. Pattern Recognition Letters, 2015, 63(1): 36-42.

[18] Jana P, Ghosh S, Bera S K, *et al.* Handwritten document image binarization: an adaptive K-means based approach [C]// Proc of Conference on Calcutta Conference Lalit: IEEE Press, 2017: 226-230.

[19] Yin Xucheng, Pei Weiyi, Zhang Jun, *et al*. Multi-orientation scene text detection with adaptive clustering [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1930-1937.

[20] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [C]//Proc of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich: Springer Press, 2015: 234-241.

[21] Gatos B, Ntirogiannis K, Pratikakis I. ICDAR 2009 document image binarization contest (DIBCO 2009) [C]// Proc of the 10th International Conference on Document Analysis and Recognition, Barcelona: IEEE Press, 2009: 1375-1382.

[22] Pratikakis I, Zagoris K, Barlas G, *et al*. ICFHR2016 handwritten document image binarization contest (H-DIBCO 2016) [C]// Proc of the 15th International Conference on Frontiers in Handwriting Recognition, Shenzhen: IEEE Press, 2016: 619-623.

[23] Pratikakis I, Zagoris K, Barlas G, *et al.* ICDAR2017 competition on document image binarization (DIBCO 2017) [C]// Proc of the 14th IAPR International Conference on Document Analysis and Recognition, Kyoto: IEEE Press, 2017: 1395-1403.